https://doi.org/10.31256/Wt3Yp1E

A Comparison Study of Neural Network-Based Semantic Segmentation for Off-Road Traversability

Semih Beycimen
Centre for Autonomous and Cyber-Physical Systems, *SATM*
Cranfield University, Bedford, UK
semih.beycimen@cranfield.ac.uk

Dmitry Ignatyev
Centre for Autonomous and Cyber-Physical Systems, *SATM*
Cranfield University, Bedford, UK
d.ignatyev@cranfield.ac.uk

Argyrios Zolotas
Centre for Autonomous and Cyber-Physical Systems, *SATM*
Cranfield University, Bedford, UK
a.zolotas@cranfield.ac.uk

***Abstract*—This paper presents work from a PhD study on unmanned ground vehicle advanced traversability. In particular, in this paper a number of learning algorithm have been trained and tested using the YAMAHA dataset (an off-road related dataset). Results were analysed and compared in terms of prediction accuracy and training time. It was noted that while various models provide appropriate accuracy results, only few provide results that can be classed as optimal when training time is considered.**

***Index Terms*—semantic segmentation, deep learning, off-road traversability, ugv, neural networks**

I. Introduction

Semantic segmentation is an important topic in many robotics related applications, and numerous studies have been published in the literature already. We list and briefly discuss important ones linking to the work presented here.

The authors in [1] proposed a multi-modal semantic segmentation method to train their model with several data augmentation methods such as *cropping*, *rotation*, *scaling* implemented to increase the dataset size.They claimed that their method provided more satisfactory resulting terms of classification accuracy and time compared to alternative methods (at the time of the publication). Another study, i.e. [2], presented a large dataset that was obtained based on weakly supervised semantic segmentation (for terrain class) and a ground reaction score. The data augmentation methods were implemented using both RGB images and labelled images with a footpath used as input for the neural network algorithms. The method was validated using an ANYmal quadruped robot in unstructured environment including terrain types such as *asphalt*, *dirt*, *sand*, *grass*, and also considering different weather and light conditions. From a more human-driven knowledge viewpoint, work in Onozuka et al [3] proposed an automatic labelling system based on human-driven knowledge. A two-step approach was followed i.e. offline training with trained images and online process performing semantic segmentation. A very recent study , i.e. [4], utilised RGB images, LiDAR point cloud and robot motion data to generate a traversability cost map. RGB images were converted to semantic mask using a supervised semantic segmentation algorithm.

In this paper, various semantic segmentation neural network (NN) model have been investigated referring to their overall accuracy, the labelled class accuracy and training time. The models have been trained based on the YAMAHA dataset, and then tested with portions of the YAMAHA dataset and our own (in-house) dataset. Our approach is presented in the methodology section, and further discussion is presented the results and discussion section.

II. Dataset used for this work

We utilise the *YAMAHA* dataset [5] to train (906 data subsets), validate (145 datasubsets) and test (80 subsets). It is worth mentioning that in the test phase 24 data subsets were used from the YAMAHA data and 56 data subsets used from the authors own data collection that was obtained using a Husky A100 platform equipped with Zed2 cameras. The dataset comprises eight classes that are: *rough trail*, *smooth trail*, *non-traversable low vegetation*, *traversable grass*, *obstacle*, *high vegetation*, *sky* and truck (background). Dataset example is shown in Figure 1.

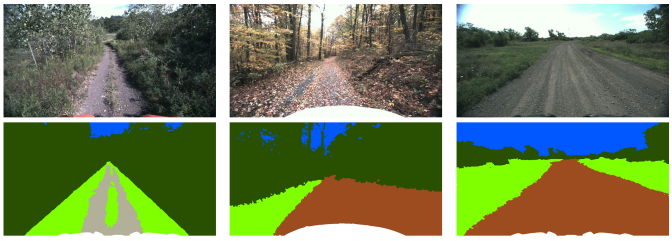

Fig. 1. YAMAHA Dataset - Above: RGB images, Below: Segmented Images

III. Methodology

We present a thorough testing and comparison of several learning segmentation models to obtain insights on the relevant performance for traversability. It is worth mentioning the use of parameters such as computational consumption, training time and prediction for real time tests for the comparison. For semantic segmentation, the following neural networks (NN) have been used, namely: $FC-DenseNet56$, $FC-DenseNet67$, $FC-DenseNet103$, $Encoder-Decoder$, $Encoder-Decoder-Skip$, $RefineNet10$, $FRRN-A$, $FRRN-B$, $MobileUNet$, $MobileUNet-Skip$, $PSPNet$, GCN, $DeepLabV3$, $DeepLabV3_plus$, $AdapNet$, $DenseASPP$, $BiSeNet$, $Modified-BiSeNet$.

TABLE I
COMPARISON OF SEMANTIC SEGMENTATION NN METHODS

Network Model	No. of Total Parameters	Training Time (hrs)	Overall Accuracy	Trvsble Grass	Rough Trail	Smooth Trail	Obstacle	Truck (Bckgnd)	Non Trvsble	Sky	High Veg.
FC-DenseNet56	1 365 632	08:22:28	79.5	72	70	82	73	72	68	97	94
FC-DenseNet67	3 442 536	08:17:54	77.6	76	69	74	76	75	68	88	93
FC-DenseNet103	9 269 784	08:30:12	78.0	59	70	84	73	76	71	96	94
Encoder-Decoder (based on SegNet)	34 968 264	07:33:22	78.8	69	67	81	73	79	74	94	93
Encoder-Decoder (with Skip)	34 968 264	07:36:48	79.3	73	74	83	72	78	65	95	95
Refinet	85 684 616	07:58:30	75.5	67	70	80	75	77	68	93	96
FRRN-A	17 740 088	10:50:48	78.2	65	65	80	74	78	69	96	93
FRRN-B	24 748 984	10:52:44	79.5	72	68	83	75	78	66	95	96
Mobile-Unet	8 872 392	07:56:05	79.2	73	69	84	73	76	67	91	95
MobileUNet-Skip	8 872 392	07:56:47	79.4	69	64	92	75	77	68	95	95
PSPNet	56 000 328	05:44:21	74.7	61	71	71	72	78	69	95	92
GCN	42 994 642	05:38:16	75.2	63	66	86	74	76	69	95	94
DeepLaB-v3	46 661 256	04:34:23	74.3	76	71	78	72	72	66	95	91
DeepLaB-v3+	47 955 128	04:43:30	78.7	72	71	75	76	79	68	97	97
AdapNet	21 069 187	05:35:19	76.5	60	65	86	73	73	70	94	92
DenseASPP	43 767 112	04:23:05	77.4	56	63	85	74	79	73	92	93
BiSeNet	47 564 712	05:45:51	78.9	64	64	88	77	78	69	96	95
Modified BiSeNet	44 476 424	06:38:28	79.1	68	70	87	79	79	70	96	96

Note: Values in columns 4-9 given in %; Trvsble = Traversable; Bckgnd = Background; Veg. = Vegetation

IV. RESULTS AND DISCUSSION

Training was performed using a super computer with V100 GPU card and two Intel E5-2620 v4 (Broadwell) CPUs (HPC). The following were used during the learning phase: batchsize 1, Root Mean Squared Propagation (RMSProp) with decay 0.995 and learning rate 0.001. We discuss results on accuracy and training time as well as labelling for the models below.

A. Overall accuracy and training time for the NN Models: The obtained results have been listed in the fist four columns of Table I for the methods discussed previous section. From the results, accuracy rates of validation data for several models provide similar performance, however once training time is taken into account it is clear that *DenseASPP, DeepLaB-v3+, BiSeNet* and *Modified BiSeNet* (with extra 2 Layer) are the highlighted methods. Note that the accuracy rate did not increase after about 30-40 epochs (models have many layers and dataset has limitations), this can be mitigated by adding more labelling data or increasing data size with data augmentation methods.

B. Accuracy results for each Label for NN models: The accuracy rate for each label class have been presented in the rest of Table I. Sky and high vegetation prediction is always high, while for traversable grass, rough trail and not-traversable vegetation is limited. Note that sky and high vegetation labels are dominant in the dataset distribution, while high vegetation, not-traversable vegetation and traversable are similar in terms of colour distribution and may not be detected by the algorithm exactly. It is worth mentioning that the number of labels impacts reliable segmentation if it is not sufficient, especially in terrain transition zones.

V. CONCLUSIONS

We presented testing and comparison of several segmentation algorithms. Noticeable difference in semantic segmentation model performance is highlighted once training time is taken in account in the comparison. To increase training accuracy, more comprehensive dataset and label classes can be used. To increase training time, a skip connections between networks (skips some layers in the NN and feeds the output of a layer as the input to the next layers) or an intelligent network model (doesn't have to be an architecture with too big layers) may be used. Results of this study inform future work looking at use of a new dataset, segmentation network and novel label classes to further enhance reliable segmentation for traversability.

VI. ACKNOWLEDGEMENT

The first author acknowledges Republic of Turkey, Ministry of National Education (YLYS), for supporting the study under PhD scholarship ref. U9BYTAB2LDGA7LK.

REFERENCES

[1] A. Valada, R. Mohan, W. Burgard, Self-Supervised Model Adaptation for Multimodal Semantic Segmentation, International Journal of Computer Vision 128 (5) (2020) 1239–1285. arXiv:1808.03833, doi:10.1007/s11263-019-01188-y.

[2] L. Wellhausen, A. Dosovitskiy, R. Ranftl, K. Walas, C. Cadena, M. Hutter, Where should i walk(Predicting terrain properties from images via self-supervised learning, IEEE Robotics and Automation Letters 4 (2) (2019) 1509–1516. doi:10.1109/LRA.2019.2895390.

[3] Y. Onozuka, R. Matsumi, M. Shino, Weakly-supervised recommended traversable area segmentation using automatically labeled images for autonomous driving in pedestrian environment with no edges, Sensors (Switzerland) 21 (2) (2021) 1–22. doi:10.3390/s21020437.

[4] T. H. Y. Leung, D. Ignatyev, A. Zolotas, Hybrid Terrain Traversability Analysis in Off-road Environments, in: 2022 8th International Conference on Automation, Robotics and Applications (ICARA 2022), 2022, [presented].

[5] D. Maturana, P.-W. Chou, M. Uenoyama, S. Scherer, Real-time semantic mapping for autonomous off-road navigation, in: Field and Service Robotics, Springer, 2018, pp. 335–350.